

mitoBench & mitoDB: Novel interactive methods for population genetics on mitochondrial DNA

Judith Neukamm^{1,3*}, Alexander Peltzer^{1,2,3*}, Wolfgang Haak², Johannes Krause^{1,2} and Kay Nieselt³

1 Institute for Archaeological Sciences, Archaeo- and Paleogenetics, University of Tuebingen, Germany.

2 Max Planck Institute for the Science of Human History, Jena, Germany.

3 Integrative Transcriptomics, Center for Bioinformatics (ZBIT), University of Tuebingen, Germany.

**These authors contributed equally to the study.*

Despite the availability of modern next generation sequencing technologies and therefore nuclear human genomes, the sequencing and analysis of mitochondrial DNA (mtDNA) is still common. Especially in the research field of ancient DNA and the context of population genetics, mtDNA is often the only proxy available to study extinct populations and their relationship with modern populations. As a consequence, many population genetic studies rely on the analysis of mtDNA.

A plethora of methods for the analysis of mtDNA exist, that address questions in population genetics, phylogeny and others. However, these tools typically rely on different file formats and often require manual interaction with the data for downstream analysis. Ultimately, these steps can be cumbersome, especially for non-bioinformaticians, resulting in an increased risk of errors during the analysis.

To tackle these issues, we present mitoBench and mitoDB. mitoBench is a workbench to interactively analyze and visualize mitochondrial genomes with a focus on population genetics. The graphical user interface is kept simple, to accommodate even users without further prior knowledge on computational methods. Furthermore, it shows additional information such as metadata and statistics. Currently, mitoBench offers automatic file conversion tools to connect the workbench with existing analysis methods such as BEAST[BHK⁺14], Arlequin[ELS05] and others. It also provides basic downstream analysis methods to investigate correlations between populations, such as principal component and Fst analysis.

MitoDB aims at providing a large reference panel of modern and ancient mitogenomes for population genetics. The current prototype provides a basis of around 1,000 complete mitogenomes from the 1000 Genomes project[Siv08] and will be further extended until publication. The whole system is developed as a free web-service that provides interactive and exploratory access to the database itself.

Our ultimate aim is to provide a central reference database for population genetics studies on complete mitogenomes that can be easily accessed both via a web interface and the accompanying mitoBench application, enabling users to perform typical analysis procedures much faster and more conveniently than before.

References

- [BHK⁺14] Remco Bouckaert, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh-Hsi Wu, Dong Xie, Marc A Suchard, Andrew Rambaut, and Alexei J Drummond. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, 10(4):e1003537, 2014.
- [ELS05] Laurent Excoffier, Guillaume Laval, and Stefan Schneider. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online*, 1:47, 2005.
- [Siv08] Nayanah Siva. 1000 Genomes project, 2008.